DOCUMENT RESUME

ED 284 429 FL 016 820

AUTHOR Axelrod, Melissa
TITLE Word-based Morphology: Some Problems from a Polysynthetic Language.
PUB DATE Dec 86
NOTE 16p.; Paper presented at the Annual Meeting of the Linguistic Society of America/American Association for Applied Linguistics (New York, NY, December 27-30, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adverbs; *Artificial Speech; *Athapascan Languages; Dialects; Language Patterns; *Linguistic Theory; Morphemes; *Morphology (Languages); Phrase Structure; Suffixes; *Syntax; Uncommonly Taught Languages; *Verbs
IDENTIFIERS *Koyukon

ABSTRACT

Some of the problems inherent in a word-based hypothesis asserting that the word/stem is taken as the minimal sign not only for syntax but also for morphology are examined in an analysis of a polysynthetic language, Koyukon, an Athabaskan language of Alaska. Data from the Central dialect is considered in the analysis. A brief sketch of the verbal system of Koyukon is presented, followed by consideration of the following specifics of the word-based hypothesis: the Binary Branching Hypothesis and the No Phrase Constraint. The remainder of the paper outlines some of the complications involved in a computational analysis of Koyukon following a theory of word-based morphology. (CB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284429

WORD-BASED MORPHOLOGY: SOME PROBLEMS FROM A POLYSYNTHETIC LANGUAGE

by Melissa Axelrod

Paper presented at the Annual Meeting of the LSA/AAAL/ADS
(New York, NY, December 27-30 1986).

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Axelrod

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 016820

BEST COPY AVAILABLE

Melissa Axelrod
University of Colorado, Boulder

Word-based Morphology: Some Problems from a Polysynthetic Language

Aronoff (1976) proposes a theory of word or stem-based morphology in which the word/stem is taken as the minimal sign not only for syntax but also for morphology. As Scalise (1985) puts it, "All regular word-formation processes are word-based. A new word is formed by applying a regular rule to a single already existing word. Both the new word and the existing one are members of major lexical categories." This word based hypothesis is claimed to be universally valid. In the following, I intend to examine some of the problems this hypothesis presents for the analysis of a polysynthetic language, Koyukon. Koyukon is an Athabaskan language of Alaska. It is a moribund language, with approximately 650 speakers left (Krauss, 1982), all over the age of 20 (Krauss, 1980). There are three main dialects of Koyukon; this paper will consider data only from the Central dialect. After a brief sketch of the verbal system of Koyukon, the following specifics of the word based hypothesis will considered: 1. The Binary Branching Hypothesis and 2. The No Phrase Constraint. The remainder of the paper will outline some of the complications involved in a computational analysis of Koyukon following a theory of word based morphology.

Structure of the Koyukon Verb

The Koyukon verb is made up of a stem plus prefixes. The stem is composed of root plus suffixes (and/or ablaut, and/or lengthening) indicating mode and aspect. Prefixes indicate person, status, gender,

mode, adverbials, and transitivity. The prefix complex of Athabaskan languages can be analyzed as a template comprised of 9 basic positions. The two positions to the immediate left of the stem are occupied by classifier elements -- a l component and a d component. The third position is the aspect/mode prefix *nə-. This prefix appears in neuter and perfective derivations. The 4th prefix position can be filled by the 1 sing. and 2 sing. and pl. subject pronouns. The 5th prefix position is that occupied by aspect, mode, and status (negative vs. affirmative) prefixes. Thematic, anatomical, and gender prefixes occupy the 6th position. The directive, conative, and future prefixes fill the 7th position. The 8th position is that region occupied by object prefixes and the so-called deictic subject prefixes (1 pl., 3 pl., areal). In the 9th and last position are incorporated nouns and bound adverbials. Some of these prefixes are obligatory in the sense that the particular position they occupy must be filled by one of the possible variants of that prefix class in every derivation of the verb. These obligatory prefixes are the person, mode, and classifier prefixes. Other prefixes, i.e., gender, adverbial, status, are optional -- the positions these prefixes occupy may or may not be filled depending on the meaning of the derivation. Some prefixes are thematic, that is, they have become lexicalized and are present in every derivation of a particular verb. These thematic prefixes along with the root of the verb are referred to as the verb theme.

Freidrich (1974:3) defines the verbal theme as "a basic variant defined both in terms of semantic features and the derivational morphology." As noted above, an Athabaskan verb theme consists of a root and the thematic prefixes occuring in all derivations. It is the underlying skeleton of the verb to which prefixes or strings of prefixes and suffixal elements are added in producing an utterance.

The theme itself has a meaning and, according to Rice (1983) is the basic unit of the Athabaskan verbal lexicon. The theme is inflected for person, mode and aspect, and may also take a gender prefix or derivational prefix string. The theme meaning 'go by boat, paddle', for example, is ɸ + qɑɬ (where ɸ is the classifier prefix and qɑɬ is the verb root). A derivational product of this theme, inflected for person, mode, and aspect is:

1) nəsqɑɬn̥ 'I arrived paddling'

nə	+	sə	+	ɸ	+	qɑɬn̥
perf	+	1 sg.	+	classifier	+	momentaneous stem

A derivational prefix string is composed of one or more prefixes which combine to add a particular meaning to the theme. Derivational strings may be aspectual, that is they may occur only with a particular mode prefix and aspect, or they may be non-aspectual, i.e., strings which appear with a verb regardless of mode and aspect. Non-aspectual derivational strings include the reflexive, causative, passive, and benefactive. Aspectual derivational strings have meanings which are generally adverbial in nature. One of this latter type of derivational string which may be added to the theme shown above (ɸ + qɑɬ 'go by boat, paddle') is ni + ɣɔ # (nə- impf., nə- perf.) + momentaneous stem '(go) ashore'. (Note that the # indicates a morpheme boundary. In Athabaskan terminology this boundary separates disjunct from conjunct prefixes. Disjunct prefixes are those to the left of the # boundary marker (in prefix position 9) and are less tightly bound phonologically than are conjunct prefixes though they remain bound prefixes.) An inflected form containing this string is:

2) niɣənəsqɑɬn̥ 'I paddled ashore'.

Note that the required mode prefixes and aspect are included as part of the string. It is this type of derivational prefix string which is

crucial in the examination of the whole aspectual system and which presents some of the most significant difficulties for a word or stem-based theory of morphology.

Aspect in Koyukon is expressed by a complex, multi-dimensional system. There are two major temporal categories which combine in expressing the temporal contour of the state or activity described by the verb. The first of these two categories is called mode and includes imperfective, perfective, future, and optative. The second category, which is called aspect, crosscuts the first and characterizes the event in terms of durativity, punctuality, cyclicity, etc. There are approximately 18 Koyukon aspects, each of which is marked by a pattern of stem variation including suffixation and/or vowel lengthening and/or ablaut. Each aspect has a particular stem shape for each of the four modes. For example, the following table illustrates the stem shapes for all four modes of several of the aspects possible for the root -qɔɛ 'go by boat, paddle':

	impf.	perf.	future	optative
momentaneous	qɔɛ̂x	qɔɛn̥	qɔɛl̥	qɔɛl̥
durative	qɔɛ	qɔɛn̥	qɔɛl̥	qɔɛl̥
persistive	qɔɛ̂x	qɔɛx	qɔɛx	qɔɛx
perambulative	qɔɛ̂x	qɔɛn̥	qɔɛk	qɔɛk
customary	qɔɛ̂x	qɔɛk	qɔɛk	qɔɛk
neuter	qɔɛn̥	qɔɛʔ	qɔɛʔ	qɔɛʔ

The Athabaskan modes are marked not only by stem shape but also by one or more prefixes. The optative is marked by a ɣu- prefix and the future by tə + ɣə prefixes in all aspects. Imperfective and perfective are marked by one of several possible prefixes depending on the aspect. For example, verbs of the momentaneous aspect may have a nə- imperfective mode prefix and nə- perfective prefix as in the

examples shown above, 1) nəsqɛn̨ 'I arrived in a boat' and 2) niɣənəsqɛn̨ # 'I paddled ashore'. Momentaneous derivations are also possible with a ∅ imperfective and ɣ- perfective prefix, a lə- imperfective and lə- perfective prefix, a ∅ imperfective and lə- perfective prefix, or a ∅ imperfective and ∅ perfective prefix. Each different combination enters into its own particular set of aspectual derivational strings. Other aspects do not allow this range of possible perfective and imperfective prefixes. Verbs in the semelfactive aspect, for example, may have only ∅ imperfective and lə- perfective mode prefixes.

Quite often an aspect is marked by more prefixes than just the combination of mode prefix and aspectual stem variant. One can think of the combination of elements added to the theme in order for the theme to be inflected for aspect as the 'zero' aspectual derivational string. There are some 300 aspectual derivational strings in Koyukon, each of which adds a particular meaning to the derivational product. Most choose (or trigger) specific imperfective and perfective mode prefixes and aspectual stem forms but many of the non-zero aspectual derivational strings are not limited to use with only one aspect. Many of the strings used with momentaneous stems can also be used with neuter stems. Sometimes a derivational prefix string can occur with any aspect that has a particular mode marker. For example, P + qɔ+ nə # 'on the surface of P' can occur with any lə- perfective derivation (e.g., conclusive, lə- momentaneous, continuative).

The Binary Branching Hypothesis

The Binary Branching Hypothesis as discussed by Scalise (1985) states that a word formation rule attaches one and only one affix at a time, i.e., the internal structure of a derived word will always be binary. Because of the pervasive nature of discontinuous dependencies

in Koyukon, it may be more insightful to think of the affixation process as similar to the superimposition of transparency diagrams one atop the other rather than as a binary or nesting structure.

As described above, aspect in Koyukon is marked by more than a single affix. In addition to the affixation/ablaut/lengthening processes by which the stem is derived from the root, a mode prefix is obligatory and, in some cases (e.g., the perambulative, certain momentaneous verbs), adverbial prefixes may also be required in the marking of aspect. In other words, the zero aspectual derivational string may contain a string of two or more discontinuous prefixes in addition to stem variation. The addition of such a string may be better described as a Gestalt than as a binary process. In many cases it is possible to find a less complex 'lexeme' onto which the derivational string can be said to be superimposed, but in general the strings cannot be broken into individual morphemes with some ordering involved in their affixation to a more basic form. Note that since the mode prefixation is dependent on stem marking of aspect and vice versa, Koyukon can only be described by a context sensitive grammar.

Non-aspectual derivations are also problematic in just this way. Examples of non-aspectual derivations will be presented with the verb root ɣɮc 'accumulate due to melting'. This root has the following stem shapes depending on mode and aspect:

	impf.	perf.	future	optative
transitional	ɣɮs	ɣɮc	ɣəsʎ	ɣɮs
perambulative	ɣɮs	ɣəsk	ɣəsk	ɣəsk
durative	ɣɮs	ɣɮs	ɣɮs	ɣɮs
customary	ɣəsk			
progressive	ɣəsʎ			

Every derivative of this verb has a nə- prefix (which can be assigned no particular meaning) and ∅ classifier prefix. These, then, are said to be thematic prefixes and the theme from which all verbs with this root are derived is described as nə + ∅ + ɣɛc '(liquid) accumulates (from melting snow, ice, or fat)'. The least morphologically complex derivatives of this theme are perfective transitional ones with ∅ perfective prefix:

3) niɣɛc 'it (e.g., water) accumulated'

A causative derivation changes the ∅ classifier to a l̥- classifier and adds the third singular object prefix yə-. The third singular subject prefix is ∅.

4) yənil̥ɣɛc 'he made it accumulate'

Note that this single derivational operation entails the simultaneous addition of two separate affixes (the direct object prefix and the l̥- classifier).

Perfective momentaneous derivations of this verb take the perfective momentaneous stem shape -ɣɛc plus one of three possible perfective prefixes. Momentaneous derivations most frequently also require one of the set of bound adverbials or adverbial prefix strings. The following example has the nə- perfective prefix plus the (non-aspectual) prefix string dɔ#də- which carries the meaning, 'a lot':

5) dɔdənəniɣɛc 'a lot of it (liquid) accumulated'

A nominal subject can be incorporated into the prefix complex in a position following the dɔ- adverbial prefix. In the following example, tu 'water' is incorporated:

6) dɔtudəmənniɣɛc 'a lot of water accumulated (from melted ice or snow, forming a pond or large puddle)'

Analyzing such a form into nesting structures would seem impossible; it is difficult to imagine a tree structure or bracketing, binary or otherwise, that could adequately describe the Koyukon verb. One could argue that it is possible to view non-aspectual derivational strings like the ones shown above as single discontinuous affixes. This possibility must be rejected though since these strings can be analyzed into separate elements which function elsewhere in the language in other combinations. The ɬ- classifier prefix which is used to mark causative, for example, also appears as a thematic classifier in many themes and is added to a theme by aspectual derivational strings such as,

7) hʉydɔ##də+ ɬ + (ɣə- impf.) + neuter stem 'be a long/short winter'.

The dɔ#də prefix string exemplified above with the meaning 'a lot of' is apparently isomorphic with another non-aspectual derivational string indicating that the activity described by the verb is a verbal one and also with an aspectual string requiring a nə- perfective prefix and momentaneous stem which carries the meaning 'all over, all up'. Aspectual derivational strings are also ill-suited to consideration as single entities since most trigger the use of one or more aspects which can occur with either a perfective or imperfective prefix.

Aspectual derivational strings present another serious problem for the Binary Branching Hypothesis: the lack of a consistent correlation between the number of affixes and the number of derivational operations. Compare the following derivatives of the verb root yɔ '(singular) go by foot':

8) niyɔ 'he/she arrived'

9) nɔ'idəyɔ 'he/she came back'

10) niniyɔ 'she walked up to a point'

11) ninɔ'idəyɔ 'she came back (to a point)'

12) hɤnɔdidəyɔ 'she started back out again'

Example (8) is clearly the least morphologically complex of these derivatives, containing only the nə- perfective marker and momentaneous perfective stem (i.e., the zero aspectual string) plus thematic ϕ classifier and third singular person inflection. In (9) the iterative derivation has applied, adding the nɔ# iterative prefix and the d element to the classifier. (10) contains the aspectual derivational string ni# (nə- impf., nə- perf.) + momentaneous stem, which carries the meaning 'going to a point, place (in time, space) and stopping.' (11) is the iterative derivative of (10). Notice that the iterative derivation, which has applied to the form produced by addition of the aspectual derivational string, prefixes elements to the right of those prefixed previously -- closer to the stem instead of further away.

(12) is also a product of an iterative derivation. In this case, the form to which the iterative has applied has been marked for aspect by the application of the following aspectual derivational string: hɤ # də+ i + (ϕ perf.) + momentaneous stem 'beginning, starting, extending out laterally'. Both (11) and (12) have undergone an equal number of derivations (aspectual, iterative) and both then, must be said to be equally morphologically complex, yet (12) has a greater number of affixes.

No Phrase Constraint

Aronoff (1983) discusses the question of "the extent to which descriptions of phenomena which lie unquestionably within the domain of word formation must make reference to syntactic notions" (369). The

traditional position, he says, is to restrict the operation of word formation rules to major lexical categories. Scalise (1985) too states that a word formation rule can take as its base only major lexical categories (N, A, V) but not phrases (NP, AP, VP) or sentences. Aronoff (1983) however, expands the domain of word formation rules to include phrasal categories in order to account for object incorporating languages and to make it possible to inclde subcategorization frames within lexical entries. Word formation rules involving verbs then, according to Aronoff, may include reference "to direct objects, indirect objects, or manner adverbs, because these are part of the verb phrase, but not to time adverbs or subjects because they are outside the verb phrase" (370).

In order to describe word formation rules involving verbs in Koyukon, however, information about material outside the subcategorization frame of a given word must also be allowed, since both subject incorporation and adverbs of time included within aspectual derivational strings are important and productive components of the language. The following data illustrate subject incorporation with both intransitive and transitive verbs.

13) yədɔkʼəqʼʁƛ̓niyɔ 'cold air (kʼəqʼʁƛ̓#) rushed in (when the door opened)'

14a) John niniyɔ 'John went up to a point (and stopped)'

b) nitɔniyɔ 'the water (tɔ#) stopped rising'

15a) John hɔlyɔ 'John went up the bank'

b) hɔtɔlyɔ 'the water (tɔ#) rose over the bank'

16) nicʼix̂niɣəl̥ 'the boat (cʼix̂#) came to a stop'

17a) aƛ̓tɔn̥ 'it (dead fish) is there'

b) John yəx yiƛ̓tɔn̥ 'John is keeping it (dead fish) in the house (yəx##)'

c) hʉt'ətɔyədilʈʐn 'water (tɔ#) (i.e., the current) is keeping it in between logs (hʉt'ə##)'

18) q'ɔ'əlc'ix̂yi'ɔx̂ 'the wind (əlc'ix̂#) is blowing it around'

Most of the adverbial aspectual derivational prefix strings in Koyukon refer to place or manner but several can be construed as adverbs of time, including the following:

19a) P+ɔ#time incorp.#hʉ + (ɣə- prog.) + progressive-momentaneous stem 'P was out, gone all of the time indicated by incorp. (day, night, summer, winter)'

b) N#(#)(də)#(ɔ)ɣɔ#k'ə + (ɣə- perf., prog.) + momentaneous or progressive stem 'all N (morning, evening, spring, fall) long'

c) ʒɔn# (ɣə- neuter, ɣə- momentaneous) 'during the day, the span of a day'

d) ʒɔn#(#)k'ə+ (ɣə- neuter, ɣə- momentaneous) 'all day long'

e) ɬəd #k'ə + (ɣə- neuter, ɣə- momentaneous) 'all night long'

f) yi#qʼə#nɔ#(k'ə) + (lə- momentaneous, ɣə- progressive) 'all night'

It seems then that the boundary between lexical category, phrasal category, and sentence is ambiguous enough in Koyukon to necessitate the inclusion of information about syntactic operations in the formulation of word formation rules.

Computational Analysis

Byrd, Klavans, Aronoff, and Anshen (1986) describe a system for computationally analyzing affixation in English. The system employs four word lists and eight computer programs which allow researchers to input morphologically complex words and receive as output a structure which identifies the base and "shows the affixes and marks the parts-of-speech of the components." To implement such a system for the

verbal system of a polysynthetic language like Koyukon, some adjustments would need to be made.

First, rather than just monomorphemic words, all roots would need to be available on a list for the programming system. Because suppletion and irregularity are so prevalent in the stem sets for each root, many roots would need to be listed with their associated stems in addition to listing the regular rules by which aspectual stem variants are derived. Further, since verbal utterances are derived from non-predictable themes, rather than from simple roots, these too would need to be entered in a dictionary. Lists of aspectual and non-aspectual derivational strings, gender prefixes, postpositions, incorporating forms of nouns, and phonological rules would be necessary as well. This would entail a certain amount of redundancy since gender prefixes and postpositions both participate in many aspectual derivational strings with unique meaning and also function as separate elements.

Because this is a language that can only be described by a context sensitive grammar, a finite state automaton would be inadequate to parse it. Parsing a verb form of Koyukon would require a device that could look both forward and backward and which could remember what it saw. It would need to operate more along the lines of Koskenniemi's TWOL, a morphological parser developed for Finnish which scans words from left to right devising possible candidates for what the word might be in the process of recognizing that word -- a process of analysis by synthesis. Such a parser for Koyukon would have to scan from right to left, first matching the stem with possible source roots, then simultaneously matching the prefix combination with possible source themes and determining the aspectual string in order to compare it with the aspectual potential of each candidate theme

(e.g., if the aspectual string is a semelfactive one, the theme from which it is derived must be one that permits semelfactive derivations). Finally, the parser would need to identify inflectional affixes.

Rather than producing a bracketed structure like the one described by Byrd et al (1986), a parser for Koyukon would need to produce the discontinuous affix strings from which the word is constructed: theme and derivational strings. Identifying the theme is a complex task. The parser would have to know, for example, the derivations which alter the classifier prefix. Passive, benefactive, reflexive, and reciprocal derivations, among others, add a d element to the classifier so that a thematic ∅ classifier will be də- in the derived form and thematic ł- will be łə- in the derived form while the thematic də- and łə- classifiers remain unchanged. The causative derivation changes any thematic classifier prefix to ł. The parser, then, must be able to see a də-, ł-, or łə- classifier and know that it might represent a thematic or derived form and proceed accordingly.

This brings us to a more important problem. If the computer scans the də- classifier prefix and finds a nɔ# prefix later, how is it to know whether it has found an iterative form, which is marked in just this way, or a verb with a thematic də- classifier and one of two other nɔ# prefixes which occur in the same position as the nɔ# iterative prefix but have entirely different meanings? Speakers are able to resolve this kind of potential ambiguity via the conversational context. Clearly, in order for the parser to make a decision of this sort, it would need access to the meaning of the word. A parser without a rather elaborate semantic component would be unable to do more than provide a list of possibilities for the identification of affixes within a given word.

REFERENCES

Aronoff, M. 1976. Word Formation in Generative Grammar. Linguistic Inquiry Monograph 1. MIT Press: Cambridge.

-----. 1983. A decade of morphology and word formation. Annual Review of Anthropology 12: 355-75.

Byrd, R.J., J.L. Klavans, M. Aronoff, and F. Anshen. 1986. Computer methods for morphological analysis. Proc. Assoc. Computational Linguistics, 120-7.

Freidrich, P. 1974. Aspect in Homeric Greek. IJAL Monograph.

Krauss, M.E. 1980. Alaska Native Languages: Past, Present, and Future. ANLC: Fairbanks.

-----. 1982. Native Peoples and Languages of Alaska (map). ANLC: Fairbanks.

Rice, K. 1983. A Grammar of Slave. Ms.

Scalise, S. 1985. Generative Morphology. Foris: Dordrecht.